



**University of  
Zurich<sup>UZH</sup>**

**Zurich Open Repository and  
Archive**

University of Zurich  
University Library  
Strickhofstrasse 39  
CH-8057 Zurich  
[www.zora.uzh.ch](http://www.zora.uzh.ch)

---

Year: 2015

---

## Computational epigenomics: challenges and opportunities

Robinson, Mark D ; Pelizzola, Mattia

DOI: <https://doi.org/10.3389/fgene.2015.00088>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-114979>

Journal Article

Published Version



The following work is licensed under a Creative Commons: Attribution 4.0 International (CC BY 4.0) License.

Originally published at:

Robinson, Mark D; Pelizzola, Mattia (2015). Computational epigenomics: challenges and opportunities. *Frontiers in Genetics*, 6:88.

DOI: <https://doi.org/10.3389/fgene.2015.00088>

# Computational epigenomics: challenges and opportunities

Mark D. Robinson<sup>1,2</sup> and Mattia Pelizzola<sup>3\*</sup>

<sup>1</sup> Institute of Molecular Life Sciences, University of Zurich, Zurich, Switzerland, <sup>2</sup> SIB Swiss Institute of Bioinformatics, University of Zurich, Zurich, Switzerland, <sup>3</sup> Computational Epigenomics, Center for Genomic Science of IIT@SEMM, Istituto Italiano di Tecnologia, Milano, Italy

**Keywords: epigenomics, histone mark, DNA methylation, ChIP-seq, epigenetic code**

The field of epigenetics is undoubtedly attracting immense interest with countless studies in various areas of investigation; (see Rivera and Ren, 2013) for a review on the state of the art for human epigenomics. From the computational point of view and the characteristics of the generated data, epigenomics is a very complex field, for two main reasons. First, epigenetics encompasses a multi-layered set of regulatory cues that act coordinately and possibly in a combinatorial way to control fundamental biological processes, such as the output of gene expression programs. Second, profiling techniques based on high-throughput sequencing are widely adopted in this field, generating comprehensive yet complex and massive genome-wide datasets. As a result, the contribution of scientists with computational skills (computer scientists, statisticians, physicists and computational biologists) is considered an essential component of research institutes investing in this research field (Bock and Lengauer, 2008).

In this Research Topic, we collected a number of contributions in the field of computational epigenomics covering three main research areas: (i) computational analyses tackling important issues closely related to the experimental method used to generate epigenetic data (Flensburg et al., 2014; Ji et al., 2014; Mensaert et al., 2015), (ii) computational approaches useful to overcome pitfalls associated to the analysis of a given epigenetic layer (Barozzi et al., 2014; Cairns et al., 2014; Robinson et al., 2014), and (iii) studies on the integration of multiple epigenetic layers (de Pretis and Pelizzola, 2014; Fejes et al., 2014; Osella et al., 2014).

Computational tools developed for the analysis of specific epigenetic data types, including DNA methylation and ChIP-seq of histone post-translational modifications (so-called “marks”), have to deal with the biases originated directly from the experimental methodology. In the case of profiling DNA methylation, various approaches based on sequencing are available, depending on the desired tradeoff between cost, coverage and data resolution. In some cases, a non-trivial subset of the DNA fragments sequenced in MBD-seq experiments, based on affinity purification through a methyl-CpG binding protein, could not be assigned to the expected reference genome. It was then shown how it is possible to assess this unanticipated proportion of unmapped reads to profile methylated viral sequences, which can be particularly relevant in certain studies (e.g., oncoviruses; Mensaert et al., 2015). On the other hand, reads from methylated DNA were shown to be over-represented in data from whole-genome bisulfite sequencing experiments. The technical reasons for this bias and the necessity of developing computational methods for correcting this issue, especially when interested in allelic methylation, were explored (Ji et al., 2014). Finally, regarding the analysis of ChIP-seq data, computational methods were shown to be helpful in clarifying how to generate reference samples necessary for the identification of enriched genomic regions. Specifically, the effect of using pull-down of the whole histone H3 or the more common input sample (whole-cell extract) were compared, showing how this choice had negligible impact on the resulting computational results (Flensburg et al., 2014).

Various computational methods have been developed for the analysis of different epigenetic data types, yet it remains difficult to understand the relative merits and performance of all the

## OPEN ACCESS

### Edited and reviewed by:

Richard D. Emes,  
University of Nottingham, UK

### \*Correspondence:

Mattia Pelizzola,  
mattia.pelizzola@iit.it

### Specialty section:

This article was submitted to  
Bioinformatics and Computational  
Biology, a section of the journal  
Frontiers in Genetics

**Received:** 18 February 2015

**Accepted:** 18 February 2015

**Published:** 05 March 2015

### Citation:

Robinson MD and Pelizzola M (2015)  
Computational epigenomics:  
challenges and opportunities.  
Front. Genet. 6:88.  
doi: 10.3389/fgene.2015.00088

available approaches. Trying to guide on the identification of the best-suited method, a number of contributions in this Research Topic focused on the comparison between computational methods and discussion to contrast the available analysis strategies. Regarding DNA methylation data, a number of methods developed for the identification of differentially methylated bases or regions were compared, while discussing the importance of experimental design, and confounders such as batch effects and cell type composition (Robinson et al., 2014); this is a very active field, evidenced by new tools emerging, such as DMRcate (Peters et al., 2015) and M3D (Mayo et al., 2014) and also highlighting the need to constantly update performance comparisons. Touching on a different data type, chromatin accessibility, various methods for the identification of footprints in DNase-seq were discussed and compared using ENCODE data (Barozzi et al., 2014). Accessibility only reveals information about presumed activity, but is commonly chosen since it is complementary to the analysis of specific epigenetic marks and provides a list of putative regulatory proteins that bind open chromatin regions. Finally, tackling the issue of the statistical modeling of read counts for ChIP-seq data, various alternatives were discussed and a method based on double Negative Binomial (i.e., Poisson distributed counts with a mixture of two gamma-distributed rates) was proposed (Cairns et al., 2014).

While the experimental methods and the computational analysis of individual data types are compared and perfected, scientists are investigating how to make connections between the various epigenetic layers that are surveyed. It is now clear that patterns of DNA methylation and histone marks are established, maintained and have effect through a machinery that is influenced by the crosstalk between these layers, and their interplay with binding of regulatory proteins, chromatin accessibility and 3D conformation. In other words, the joint analysis of multiple epigenetic layers through data integration methods (Ritchie et al., 2015) is considered the key to comprehend how

epigenetic information contributes controlling complex regulatory processes. In this series of articles, computational and experimental methods for the integrative analysis of epigenetic marks are discussed and proposed. Double-negative feedback loops, where a microRNA is inhibited by an epigenetic regulator while being epigenetically controlled by the same regulator, are considered and shown to exhibit properties that are well suited for circuits involved in cell fate transitions (Osella et al., 2014). In the context of data integration and visualization, an online platform (DaVIE) was developed based on a database of DNA methylation experiments. This tool allows navigating through multiple DNA methylation experiments and integrating different data types, including ChIP-seq data (Fejes et al., 2014). Finally, recent and past evidence in favor of the notion of epigenetic code are discussed, and computational and experimental strategies are proposed that can be instrumental to further investigating how different epigenetic layers and marks are interconnected (de Pretis and Pelizzola, 2014).

Altogether, this series of articles provides a comprehensive glance at the emerging field of computational epigenomics. This research area brings to the field of epigenetics a set of tools that were initially developed in the field of genomics. At the same time, computational epigenetics is showing its maturity toward closing the circle between the genome and the epigenome, revealing how regulatory layers are interconnected and highlighting the need to jointly consider epigenetic phenomenon to explain complex transcriptional responses.

## Acknowledgments

MDR acknowledges financial support from SNSF (grant 143883) and from the European Commission's RADIANT project (Grant Number: 305626). MP acknowledges financial support from the European Commission's RADIANT project (Grant Number: 305626).

## References

- Barozzi, I., Bora, P., and Morelli, M. J. (2014). Comparative evaluation of DNase-seq footprint identification strategies. *Front. Genet.* 5:278. doi: 10.3389/fgene.2014.00278
- Bock, C., and Lengauer, T. (2008). Computational epigenetics. *Bioinformatics* 24, 1–10. doi: 10.1093/bioinformatics/btm546
- Cairns, J., Lynch, A. G., and Tavaré, S. (2014). Quantifying the impact of inter-site heterogeneity on the distribution of ChIP-seq data. *Front. Genet.* 5:399. doi: 10.3389/fgene.2014.00399
- de Pretis, S., and Pelizzola, M. (2014). Computational and experimental methods to decipher the epigenetic code. *Front. Genet.* 5:335. doi: 10.3389/fgene.2014.00335
- Fejes, A. P., Jones, M. J., and Kobor, M. S. (2014). DaVIE: database for the visualization and integration of epigenetic data. *Front. Genet.* 5:325. doi: 10.3389/fgene.2014.00325
- Flensburg, C., Kinkel, S. A., Keniry, A., Blewitt, M. E., and Oshlack, A. (2014). A comparison of control samples for ChIP-seq of histone modifications. *Front. Genet.* 5:329. doi: 10.3389/fgene.2014.00329
- Ji, L., Sasaki, T., Sun, X., Ma, P., Lewis, Z. A., and Schmitz, R. J. (2014). Methylated DNA is over-represented in whole-genome bisulfite sequencing data. *Front. Genet.* 5:341. doi: 10.3389/fgene.2014.00341
- Mayo, T. R., Schweikert, G., and Sanguinetti, G. (2014). M3D: a kernel-based test for spatially correlated changes in methylation profiles. *Bioinformatics*. doi: 10.1093/bioinformatics/btu749. [Epub ahead of print].
- Mensaert, K., VanCrickinge, W., Thas, O., Schuurings, E., Steenbergen, R. D. M., Wisman, G. B. A., et al. (2015). Mining for viral fragments in methylation enriched sequencing data. *Front. Genet.* 6:16. doi: 10.3389/fgene.2015.00016
- Osella, M., Riba, A., Testori, A., Corà, D., and Caselle, M. (2014). Interplay of microRNA and epigenetic regulation in the human regulatory network. *Front. Genet.* 5:345. doi: 10.3389/fgene.2014.00345
- Peters, T. J., Buckley, M. J., Statham, A. L., Pidsley, R., Samaras, K., Lord, R. V., et al. (2015). De novo identification of differentially methylated regions in the human genome. *Epigenet. Chromatin* 8:6. doi: 10.1186/1756-8935-8-6
- Ritchie, M. D., Holzinger, E. R., Li, R., Pendergrass, S. A., and Kim, D. (2015). Methods of integrating data to uncover genotype-phenotype interactions. *Nat. Rev. Genet.* 16, 85–97. doi: 10.1038/nrg3868
- Rivera, C. M., and Ren, B. (2013). Mapping human epigenomes. *Cell* 155, 39–55. doi: 10.1016/j.cell.2013.09.011

Robinson, M. D., Kahraman, A., Law, C. W., Lindsay, H., Nowicka, M., Weber, L. M., et al. (2014). Statistical methods for detecting differentially methylated loci and regions. *Front. Genet.* 5:324. doi: 10.3389/fgene.2014.00324

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Robinson and Pelizzola. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.